

CLIMATOLOGICAL STUDIES OF  
TEMPERATURE, PRECIPITATION AND  
FOG — PARTICULARLY IN SWEDEN

INAUGURAL DISSERTATION

BY

TAGE ANDERSSON

FIL. LIC., SMÅL.

By due permission of the Faculty of Science  
of the University of Uppsala,  
to be publicly discussed in lecture hall V  
on May 8, 1970, at 10 a.m.,  
for the degree of Doctor of Philosophy

UPPSALA 1970



# CLIMATOLOGICAL STUDIES OF TEMPERATURE, PRECIPITATION AND FOG — PARTICULARLY IN SWEDEN

The thesis includes the following papers:

- I On the representativity of rain measurements. Meteorologiska Institutionen, Uppsala Universitet, Reports Nr 7, 1968.
- II Contamination of rain caused by washout from the low layers of the atmosphere. Tellus, Vol. 21, Nr 5, 1969. Also in Meddelanden från Uppsala Universitets Meteorologiska Institution, Nr 101, 1969.
- III Annual and diurnal variation of fog. Meddelanden från Uppsala Universitets Meteorologiska Institution, Nr 102, 1969.
- IV The diurnal variation of precipitation in Sweden. Geografiska Annaler, Vol. 51, Serie A, Nr 3, 1969. Also in Meddelanden från Uppsala Universitets Meteorologiska Institution, Nr 100, 1969.
- V The diurnal variation of precipitation frequency over the Weather Ship M. Journal of Applied Meteorology, Vol. 8, Nr 1, 1970.
- VI Stockholmstraktens nederbörd. Meteorologiska Institutionen, Uppsala Universitet, Reports Nr 14, 1969.
- VII Temperature and precipitation records in Sweden since the 1850's. Part A: Observational methods and error sources. Meteorologiska Institutionen, Uppsala Universitet, Reports Nr 13, 1969.
- VIII Some characteristics of temperature and precipitation in Sweden. Meddelanden från Uppsala Universitets Meteorologiska Institution, Nr 103, 1969.
- IX Principal component analyses of temperature in Sweden. A preliminary report. Appendix 2 of this paper.

(References given in I - VIII will not be repeated here.)

The present collection of papers has one "common divisor" namely treating climatological problems that roughly can be divided into two types.

The first type concerns problems which earlier have been fairly extensively treated. In such cases methods hitherto seldom used in climatology have been applied (e.g. principal component analysis) in an effort to test earlier theories or discover new facts. Papers III, VI and VIII belong to this type.

The second type deals with problems that have only been sparsely treated. Sometimes they are simple problems, which to an amazingly high degree have escaped attention, the reason probably being that climatology for a long time has been and still often is regarded as an inferior part of meteorology. "Climatology is to sum 31, 30, 28, or in complicated cases 29 numbers, then dividing by the same number in order to get the monthly mean", is an old saying which one still hears now and then. One hundred years ago, when the meteorological networks were new, these operations gave very important results, and they still do. In fact means still are prominent climatological parameters. For one thing: they form the backbone of most climate classifications. But alone they are not sufficient to describe the climate. Nowadays when so much more data, as well as electronic computers, are available several meaningful statistics may be computed. When mapped or otherwise suitably treated, they give a much better picture of the climate than the mean alone. Computing good estimates of these statistics often needs long periods, but most meteorological records cover only a few years and, when available, old data are often neglected because they are difficult to obtain or regarded as too bad. But they are not to be scorned. The methods used may seem primitive, the errors may be large and impossible to estimate, but nevertheless, used with caution, they may give good results and be necessary in order to get enough data for estimating important statistics.

In recent years principal component analysis has been applied in several fields and often proved itself a powerful tool. It has also been used in some of the works in this thesis. The technique is yet not widely known and therefore a description of it is given in Appendix 1.

One reason for the heterogeneity of this collection of papers is that at first I was not aiming at a dissertation. I was simply treating problems which occurred during my cooperation with Prof. Tor Bergeron in "Project Pluvius", or during my work as lecturer in meteorology. Aid from the Meteorological Institute and grants from the Swedish Natural Science Council and Statens Råd för Byggnadsforskning then made it possible for me to pursue these studies.

### Summaries of the papers

Papers I and II treat data from "Project Pluvius", an investigation conducted by Bergeron since 1953, aiming at a detailed description and physical explanation of the precipitation conditions within small areas, with special emphasis on studying the influence of the terrain on the small-scale precipitation distribution, i.e. estimating the influence of height above sea-level and of forest on the precipitation.

#### I. On the representativity of rain measurements

This paper tries to estimate the representativity of rain measurements, i.e. how accurately the total rainfall of an area may be estimated from one (or only a few) point measurement(s). The study is performed both for daily and monthly rainfall sums. The series of data available are fairly short, especially as to monthly sums. The annual variation has not been eliminated in the regression analysis. Therefore, the confidence intervals given in Fig. 3 may be somewhat too narrow, but probably the error is small. One of the most important items in the hydrological cycle is the precipitation; good estimates of its areal depth are necessary for water balance studies.

#### II Contamination of rain caused by washout from the low layers of the atmosphere

The intention was to get a detailed picture of the contamination of rain within a small area. Therefore a network of 24 monthly and 5 daily samplers for rainfall within an area of about 400 km<sup>2</sup> centred around Uppsala was established in June 1962. This network was much denser than any previous one

in Sweden and was working during four months. The results from the monthly samplers are treated in paper II and show that the representativity of measurements of this type is extremely low. The reason is the washout of pollutants from the lower atmosphere by the precipitation particles. Since town air is much more polluted than that of rural areas, the contamination of the rain is much higher in towns. Laboratory experiments performed by Bielke and Georgii (1968) confirm this result, as well as measurements of the contamination of snow carried out by H.Rodhe (1969).

### III Annual and diurnal variation of fog

The climatology of fog, mainly in Sweden, is treated, but in order to get a wider and deeper insight into the fog climate, observations from the surrounding seas as well as from the North Sea and North Atlantic are included. The extremely local character of fog over land makes attempts at mapping its duration or frequency a very difficult (or hopeless) task. Therefore, the approach used is to describe its annual variation with harmonic analysis. Sine waves having periods of 12, 6, 4, 3, 2.4 and 2 months were used to describe the observed annual march of the monthly mean frequencies. The underlying idea is that even if the fog frequency itself is too irrepresentative to permit a mapping, it may be possible to map its annual march, and the harmonic analysis may be a proper tool for this purpose. Over Sweden and its surrounding seas as well as over the North Atlantic the harmonic analysis proved to be fairly efficient, and the two first harmonics (period 12 and 6 months resp.) were sufficient to give a good description of the observations. Over the North Sea the harmonic analysis was inefficient. There a principal component analysis of the type used in Papers IV and VI may be the proper tool.

The different fog climates over the Swedish mainland and the surrounding seas south of  $61^{\circ}\text{N}$  are clearly described by the time of maximum for the first harmonic, which for this region "explains" most of the observed variance. At the coast of the Gulf of Bothnia a semi-annual type of variation dominates. Since the time between the maxima is about 6 months, it is efficiently described by the second harmonic.

The diurnal variation of fog, with maximum frequency at sunrise, is very pronounced over Sweden except during the coldest months. Over the oceans,

the diurnal variation is very small, and over major inland water bodies as the Baltic a diurnal variation exists similar to but much smaller than that over land. Also as to the annual variation the Baltic is intermediate between that of land and ocean, spring being the time for maximum frequency over the central Baltic. Over middle latitudes the ocean fog maximum occurs during summer and the continental one during late autumn - early winter.

#### IV The diurnal variation of precipitation in Sweden

Mainly the "classical" types of the diurnal variation of precipitation occur over Sweden. During summer the precipitation amounts have a pronounced daytime maximum over land and daytime minimum over the seas, and during winter the diurnal variation is smaller, mainly with a day deficit over land. The north-western mountain range, however, forms an exception, having a pronounced daytime maximum during winter. These features were clearly shown by a principal component analysis. A short description of the method will be given in Appendix 1. A more detailed but still fairly elementary derivation is given by Morrison, 1967. This technique, which was suggested already in 1901 by Pearson and further developed in the 1930's, was first introduced in meteorology by Lorenz, 1956.

The winter daytime maximum has evidently not been described earlier, but though its reality can hardly be doubted, no explanation can be given.

#### V The diurnal variation of precipitation frequency over the Weather Ship M

Using the 3-hr synoptic observations during 1949 - 68 this paper treats the diurnal variation of precipitation frequency over the Weather Ship M (Polarfront) at  $66^{\circ}\text{N}$ ,  $02^{\circ}\text{E}$ . The "classical" diurnal variation, viz. highest frequencies during night, prevails during the warmer months but not during winter. During January and February the lowest frequencies occur during night. The daytime excess of precipitation frequency during these months is, however, smaller than the corresponding daytime excess of precipitation amounts over the mountains of north-western Sweden.

## VI Stockholmstraktens nederbörd

This paper (in Swedish) treats the precipitation climate of the Stockholm region. It is mainly intended to be a handbook and therefore contains many tables, extracted from different sources, giving several statistics. Beside the ordinary maps of average precipitation it also contains maps giving the geographical distribution of rainfall resp. snow-depth after some severe storms, in order to illustrate the great areal variation characteristic of precipitation. The variation of precipitation sums after individual storms over a small area was discussed in paper I. Maps giving total rainfall for several months over an area of c. 500 km<sup>2</sup> surrounding Uppsala are also included. These data originate from "Project Pluvius".

An important advance in precipitation climatology is Bergeron's discovery and explanation of the coastal orographic precipitation maxima during autumn and winter. In order to test this theory, principal component analysis was performed using the mean monthly sums from c. 100 stations during the period 1950 - 63. This analysis shows that the most prominent feature of the geographical distribution of precipitation is the Bergeron coastal maximum. Even in the summer this maximum is discernible, though then largely being overshadowed by convective precipitation characteristic of the inland.

## VII Temperature and precipitation records in Sweden since the 1850's

### Part A: Observational methods and error sources

Paper VII is the first part of a work which will summarize the long-time records of monthly temperature and precipitation in Sweden; its present part describes the observational methods used and discusses their error sources. The statistics are not included, but will be published later, and the observations are available on magnetic tape.

The foundation for the climatology of Sweden is formed by two networks of meteorological stations. The first one, created in the late 1840's by the Naval Authorities, was initiated by A.Erdmann and F.Kreuger and consisted of about 20 stations at lighthouses. Temperature records from some of them are available since 1854. Already earlier meteorological observations had been performed at some lighthouses\*. A network of about the same number of land stations was established about 1860 by the Royal Swedish Academy of Science, whose physicist at that time was E.Edlund. The first reports treating results from these stations appeared in the 1870's in papers by

\* According to personal communication with Dr. C.J. Östman.

Edlund and Rubenson. By that time H.E.Hamberg had started a career devoted to the science of climatology. Hamberg established several new stations, and in many ways tried to keep their standard as high as possible and to introduce improvements. A.o. questionnaires were sent to the stations nearly every year, asking the observers to describe their sites, check their thermometers etc. Hamberg's capacity also enabled him to become the founder of Swedish climatology. His maps of temperature, precipitation etc. reveal an ardent scientist with a real understanding of the problems confronting him. His contemporaries, H.H.Hildebrandsson and N.Ekholm also made important contributions, but in climatology they cannot be compared to Hamberg, though the large stock of data on different phenomena as freezing days, thunder etc. (available at the archive of the Meteorological Institute, Uppsala) collected by Hildebrandsson really is impressive.

#### VIII Some characteristics of temperature and precipitation in Sweden

This paper is a synthesis of some statistics to be published as Part B of paper VII. Paper VIII gives a description and discussion of the annual march and geographical distribution of several parameters that are necessary for a good understanding of the climate, but hitherto have mostly been neglected.

Most of the works on the temperature and precipitation conditions in Sweden performed after the Hamberg era are carried out by persons specialized in other fields of meteorology. Amongst these A.Ångström deserves special mention for his research on the temperature conditions and his book "Sveriges Klimat" (Climate of Sweden) which for a long time has been a standard work. The standard work before Ångström's book was "Climate of Sweden" by A.Wallén; although originally being a hydrologist he had carried out some climatological studies. C.C.Wallén's studies of precipitation are also important. Bergeron's precipitation studies have already been mentioned, but perhaps the paper "Richtlinien einer dynamische Klimatologie" (1930) forms his most significant contribution because of its pleading for introduction of frequencies and intensities of prominent elements as well as physical concepts in climatology. H.Modén's comparisons between temperature observations in different screens and shelters are instructive, because in Sweden very little attention has been paid to the exposure of thermometers. Around 1900 much work was performed on finding formulas for computing monthly temperatures from three (or five) daily readings as accurately as

possible, and the much more important question of exposure seems to have been neglected. — G.H.Liljequist's studies of the winter and summer temperature from the long Stockholm record constitute the most important studies of time series. His textbook "Klimatologi" will probably become a standard work. Lack of economic resources, and a general feeling that it is not possible to get much out of the climatological records, may be the reason for the comparatively small interest in climatology. True, a very good general survey of the geographical distribution of averages was already reached by Hamberg. But, as to other parameters needed for a description of the climate, little is published. C.C.Wallén has studied the frequency distribution of annual and daily precipitation sums and mapped some measures of them. As to such an important parameter as the standard deviation of the frequency distribution of monthly temperatures, a mapping of them is performed in an unpublished manuscript by B.Johnsson, who also made a thorough study of the qualities of the different formulas used for computing the monthly temperature.

Studies of the frequency distribution of the monthly temperatures may be fruitful. In paper VIII it is shown that the geographical distribution of the coefficient of skewness is an important climatological parameter. For a good description of the temperature climate this coefficient, as well as a measure of the scatter, e.g. the coefficient of intersequential variability, are needed. Both these coefficients also have a very marked annual march.

The autocorrelation coefficient of monthly precipitation (serial correlation with lag one) is very close to zero. A much larger autocorrelation exists between the monthly temperatures, reaching its highest values over the Baltic and the Gulf of Bothnia, where coefficients as high as 0.7 exist.

The space correlations of monthly temperature and precipitation in Sweden show several interesting features. As to precipitation the space correlations are much lower than for instance in Norway where a larger proportion of the precipitation is orographic (orogenic) and thus more dependent upon the large-scale wind system than in Sweden. As to the geographical distribution of the space correlations the isopleths show a tendency to be elongated in the direction of the peninsula.

For the temperatures the space correlations are much higher and the isopleths show a marked tendency to follow the east coast. On a climatological map a coast is characterized by a bundle of isopleths for different elements, e.g. temperature, rainfall. But for temperature at least the gradients of the isopleths for the space correlation are also large. This tells us that there is not only a large difference in average temperature between nearby sea and land stations, but the actual difference varies greatly from year to year. For two nearby land stations (or two nearby sea stations) the actual temperature difference varies much less from year to year.

The analyses of space correlations may be considered as pilot studies for principal component analyses. In Appendix 2 the temperature conditions in Sweden are studied using this technique.

The coastal concentration of isopleths marking the difference between land and sea climate occurs thus not only for such simple statistics as measures of central tendency (mean, median) but also for more complicated statistics as coefficients of skewness, of autocorrelation, of space correlation etc.

# An outline of the technique of principal component analysis

## Introduction

The analyses using principal components (or empirical orthogonal functions) in papers IV and VI show that in climatological contexts this technique may be very efficient in reducing the number of variates necessary for a satisfactory description of the conditions.

Let us assume that our original data consist of monthly temperatures, e.g. January ones, from  $p$  stations during  $N$  years. For each year we can map and analyze the geographical distribution of the temperature anomalies. We thus get  $N$  maps, each containing  $p$  numbers. Altogether we thus work with  $N \cdot p$  numbers. In the component analysis we seek for a linear transformation of the type:

$$m_{ik} = \sum_{j=1}^p e_{jk} x_{ij} \quad 1$$

$x_{ij}$  = temperature anomaly (= deviation from the climatological mean) for station  $j$  during year  $i$

$e_{jk}$  = component  $j$  of eigenvector  $k$  (each eigenvector or characteristic vector consists of  $p$  components; one component belongs to each station. Together there are  $p$  eigenvectors)

$m_{ik}$  = principal component or multiplier  $k$  for year  $i$  (each year has  $p$  principal components)

The original observations are recovered by

$$x_{ij} = \sum_{k=1}^p e_{jk} m_{ik} \quad 2$$

The point is that it may be possible to get a good approximation of the  $x_{ij}$  using fewer than  $p$  of the eigenvectors and multipliers, i.e. we approximate;

$$\hat{x}_{ij} = \sum_{k=1}^q e_{jk} m_{ik} \quad q < p$$

Instead of  $N$  maps for depicting the temperature conditions we need  $q$  maps showing the distribution of  $e_{j1}, e_{j2}, \dots, e_{jq}$ , and  $q$  numbers for each year ( $m_{i1}, m_{i2}, \dots, m_{iq}$ ). In Appendix 2 it is shown that only  $q = 2$ , i.e. two maps are necessary for a good description of the anomalies of the

monthly temperature distribution during e.g. January in Sweden. These two maps give the characteristic features or patterns of the geographical distribution; the two numbers (multipliers or principal components) needed for each year may be used for an objective classification.

Using a geometrical approach some more details of the principal component analysis will be given in the next section, but for a more complete description the reader is referred to Morrison (1967).

### Some details of the principal component analysis

Assume that we have monthly temperatures from January during  $N$  years ( $i = 1, 2, \dots, N$ ) from  $p = 3$  stations ( $j = 1, 2, 3$ ). For each station we compute its arithmetic mean for January over the years (its climatological January mean) and work with the deviations from this mean,  $x_{ij}$ . (Note that in papers IV and VI we have worked with the actual values, not with the deviations from the mean.) Let us plot the  $x_{ij}$  in a three-dimensional orthogonal coordinate system.

Fig. 1 shows the original  $X$  axes and the principal axes,  $M$ . The major principal axis,  $M_1$ , has the property that the sum of squares of the projections of the observations on it has a maximum, i.e. the sums of  $(OP')^2$  in Fig. 2 have a maximum. The major principal axis passes through origo and its direction is completely determined by the direction cosines, i.e. for  $M_1$

$$e_{11} = \cos \alpha_1 \qquad e_{21} = \cos \alpha_2 \qquad e_{31} = \cos \alpha_3$$

The direction cosines are the components of the eigenvector or characteristic vector of  $M_1$ , and  $e_{11}^2 + e_{21}^2 + e_{31}^2 = 1$

The value of the observation  $(x_{i1}, x_{i2}, x_{i3})$  on the new coordinate axis  $M_1$  is

$$m_{i1} = e_{11}x_{i1} + e_{21}x_{i2} + e_{31}x_{i3} = \sum_{j=1}^3 e_{j1}x_{ij}$$

The names principal component and multiplier are used for the new coordinate,  $m$ .

The mean of the  $M_1$  variate (the first principal component or the first multiplier) is

$$\bar{m}_1 = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^3 e_{j1} x_{ij} = 0$$

The variance of the first principal component is

$$s_{m_1}^2 = \frac{1}{N-1} \sum_{i=1}^N m_{i1}^2 = \frac{1}{N-1} \sum_{i=1}^N \left( \sum_{j=1}^3 e_{j1} x_{ij} \right)^2$$

We shall choose the angles of  $M_1$  so that  $s_{m_1}^2$  will become maximized, see Fig. 2, i.e. the sum of squares of the projections  $OP'$  will have a maximum, or equivalently the sum of squares of  $PP'$  will have a minimum.

It may be shown that  $s_{m_1}^2 = \lambda_1$  where  $\lambda_1$  is known as the characteristic root or eigenvalue belonging to the direction cosines of  $M_1$ .

In the same manner we compute the direction angles of the new axis  $M_2$  so that the corresponding variance  $s_{m_2}^2 = \lambda_2$  becomes maximized. It may be shown that  $M_2$  is perpendicular to  $M_1$ . The same procedure is repeated for the third new axis,  $M_3$ . This axis is perpendicular to  $M_1$  and  $M_2$ . We need not confine ourselves to three variates, but perform this procedure for the general case of  $p$  variates in a  $p$ -dimensional space.

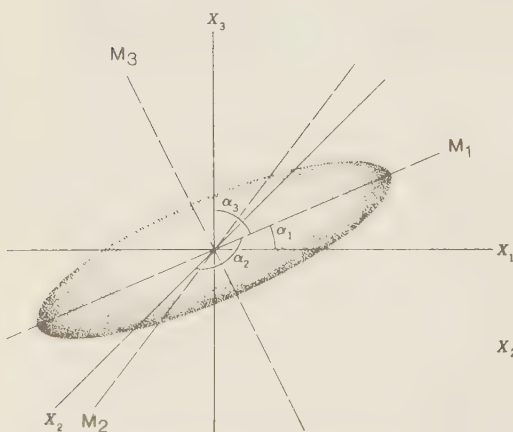


Fig. 1: Principal axes of trivariate observations (after Morrison, 1967).

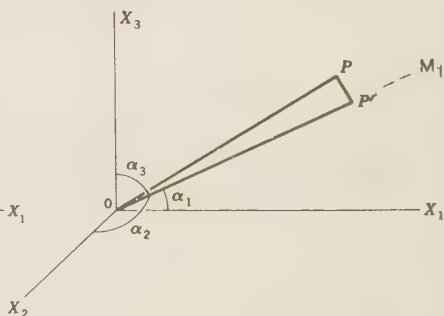


Fig. 2: Observation point  $P$  and its projection  $P'$  on the first (major) principal axis (after Morrison, 1967).

In matrix notations Eq. (1) is

$$M = E'X \quad 3$$

$M$  = the matrix of the principal components with  $p$  rows and  $N$  columns

$X$  = " " " " original anomalies "  $p$  " "  $N$  "

$E$  = " normalized orthogonal matrix of the eigenvectors

$E'$  = " transpose of  $E$ , i.e. the matrix obtained by interchanging rows and columns, row one becoming column one etc.

If the inverse matrix  $E^{-1}$  exists we have

$$EE^{-1} = I \quad 4$$

where  $I$  is the identity matrix with the diagonal elements = 1 and off-diagonal elements = 0, i.e.

$$\sum_{j=1}^p e_{jr}e_{js} \begin{cases} = 1 & \text{if } r = s \\ = 0 & \text{if } r \neq s \end{cases}$$

Since the eigenvectors are orthogonal

$$E^{-1} = E'$$

Premultiplying Eq. (3) with  $E$  gives

$$X = EM \quad 5$$

which is Eq. (2) in matrix form.

Taking the transpose of Eq. (3) gives

$$M' = X'E \quad 6$$

Premultiplying with the corresponding sides of Eq. (3) gives

$$MM' = E'XX'E \quad 7$$

This is the problem of determining the eigenvalues and eigenvectors of the symmetric matrix  $XX'$ . It may be shown that a solution exists in which the eigenvalues,  $\lambda$ , (the diagonal elements of  $MM'$ ) are real and positive, cfr Eq.(9), and the eigenvectors ( the columns of  $E$  ) are real. The off-diagonal elements of  $MM'$  are zero, see Eq. (8).

We have not worked with  $XX'$  but with the covariance matrix  $\frac{1}{N-1} XX'$ .

Its diagonal elements are the time-variances of the stations and its off-diagonal elements are the time covariances between all pairs of stations (space covariances with the terminology of paper VIII). Since we have used the covariance matrix the factor  $(N-1)$  appears in Eq. (9).

An important quality of the new coordinates,  $m$ , is that they are uncorrelated i.e.

$$\sum_{i=1}^N m_{ir} m_{is} = 0 \quad r \neq s \quad 8$$

but

$$\sum_{i=1}^N m_{ik} m_{ik} = \sum_{i=1}^N (m_{ik})^2 = \lambda_k (N-1) \quad 9$$

It may be shown that

$$s^2 = \frac{1}{N-1} \sum_{j=1}^p \sum_{i=1}^N x_{ij}^2 = \sum_{k=1}^p \lambda_k \quad 10$$

i.e. the sum of the eigenvalues equals the total sample variance.

The analysis is performed in the somewhat sanguine hope that the first few  $\lambda_k$ , let us say  $q$ , will be much larger than the remaining ones. In this case the  $q$  new coordinates account for most of the total sample variance, and the remaining  $(p-q)$  can be neglected.

Our approximation to the original data is  $\hat{x}_{ij}$

$$\hat{x}_{ij} = \sum_{k=1}^q e_{jk} m_{ik}$$

while an exact description needs

$$x_{ij} = \sum_{k=1}^p e_{jk} m_{ik}$$

$k = 1, 2, 3, \dots, p$  denotes the number of the principal axis. In words we may say that for every principal axis the direction cosines are computed. For each axis one direction cosine belongs to every station. The geographic distributions of these direction cosines may be mapped, and the analyses give, with Grimmer's (1963) terminology, the characteristic patterns, the first one referring to the major axis. To get the first approximation for the observed January temperature anomaly at station  $j$  during year  $i$  we simply compute  $\hat{x}_{ij} = e_{j1} m_{i1}$ .

For the axis belonging to the next largest eigenvalue  $\lambda_2$  we can get the second characteristic pattern in the same manner as the first one. Our improved approximation is then

$$\hat{x}_{ij} = \sum_{k=1}^2 e_{jk} m_{ik} = e_{j1} m_{i1} + e_{j2} m_{i2}$$

Proceeding in this way we may get  $p$  approximations. But if more than a few principal components are needed for a satisfactory description; thus, to account for most of the total sample variance, little is gained.

Provided that  $x_{ij}$  is the deviation from the individual mean the total sample variance is given by Eq. (10). But the analysis can be performed with other types of data, e.g. the temperatures instead of the anomalies. In such a case the mean sum of squares given by Eq. (10) is not the total sample variance.

The effectiveness of the method is usually measured by

$$\sum_k V_k = \frac{\sum_{k=1}^q \lambda_k}{\sum_{k=1}^p \lambda_k} \cdot 100$$

The method is efficient if a high  $\sum V_k$  is attained with the first few eigenvalues and the subsequent  $V_k$  drop rapidly towards zero. But even a low eigenvalue may represent a physical process. Also, due to random errors, an eigenvalue may be low only in our special sample. On the other hand, an eigenvalue may be important only in our special sample. But an eigenvalue having a high  $V_k$  probably represents a real feature though one may be unable to give a physical explanation. Even if the eigenvalue lacks one physical explanation, i.e. is the result of an interplay between several physical processes, which are hard or impossible to disentangle, the description given by it may be valuable.

It can be shown that, among all linear functions of the data, this representation is optimum in the sense that the patterns account, successively, for the largest possible fractions of the original total variance.

A comprehensive geometrical definition of principal components is given by Morrison (1967): "The principal components of the sample of  $N$   $p$ -dimensional observations are the new variates specified by the axes of a rigid rotation of the original response coordinate system into an orientation corresponding to the directions of maximum variance in the sample scatter configuration. The direction cosines of the new axes are the normalized characteristic vectors corresponding to the successively smaller characteristic roots of the sample covariance matrix. If two or more roots are equal, the directions of the associated axes are not unique and may be chosen in an infinity of orthogonal positions." Cfr Figs 1 and 2.

The normalized characteristic vectors are also called eigenvectors and the characteristic roots are also called eigenvalues.

## PRINCIPAL COMPONENT ANALYSES OF TEMPERATURE IN SWEDEN. A preliminary report

A. The characteristic patterns of the geographical distribution of monthly temperature anomalies in Sweden

The analyses of the geographical distribution of the space correlations in paper VIII may be regarded as pilot studies for principal component analyses. These will probably be most fruitful for the temperature, since this element has a much more orderly distribution than the precipitation.

The material used for this study was monthly temperatures from 31 ( $p = 31$ ) stations fairly evenly distributed over Sweden, with complete records during 1901 - 1968 ( $N = 68$ ). The analyses were performed separately for every month. There are 31 characteristic patterns for each month, hence a total of 372. However it will be shown that most of them give such small contributions that they may be neglected, and furthermore the two most important patterns nearly exactly repeat themselves every month. The first two characteristic patterns for January are shown by Fig. 3.

The first characteristic pattern has  $V_1 = 81$ , hence 81 % of the total sample variance is reproduced by it. The second characteristic pattern has  $V_2 = 12$ , and hence together they reproduce 93 % of the original variance. As an example of how the approximation works, Fig. 4 gives the approximation using only these two patterns and the actually observed anomalies for January 1913. For this special year the temperature conditions were rather complicated and the two patterns together represented only 77% of the observed variance, thus often a better representation is obtained with our approximation.

Table 1 shows  $V_k$  for the first five patterns for every month. The first pattern always accounts for about 80% of the total sample variance, the second for about 10%, while the contribution of the third never exceeds 5% and if we content ourselves with reproducing c. 90% of the total variance the two first patterns will be enough, though the third probably represents a real feature, at least during the months of April to August when for each month the isopleths for the components of the characteristic vectors are concentrated at the east coast. Cfr the coastal concentration of isopleths for the space correlation of temperature during these months demonstrated in paper VIII.

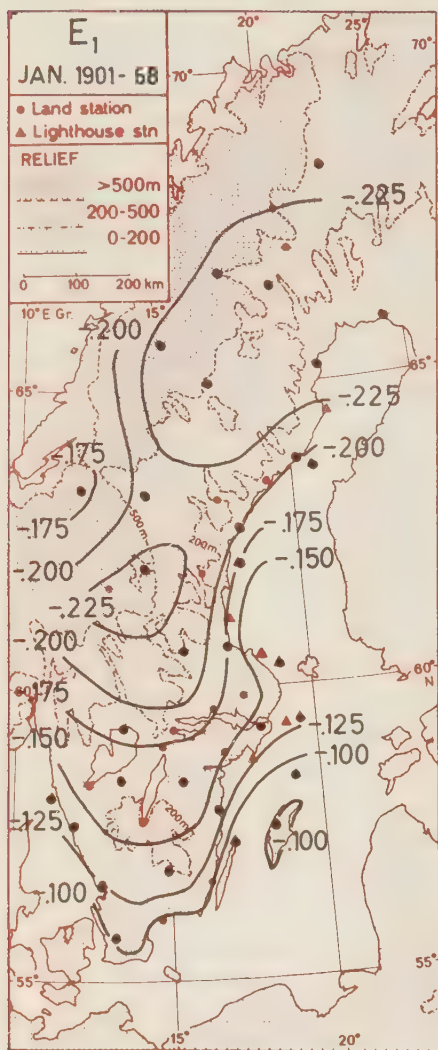


Fig. 3a: Geographical distribution of the components of the first eigenvector (the first Characteristic Pattern) for January.

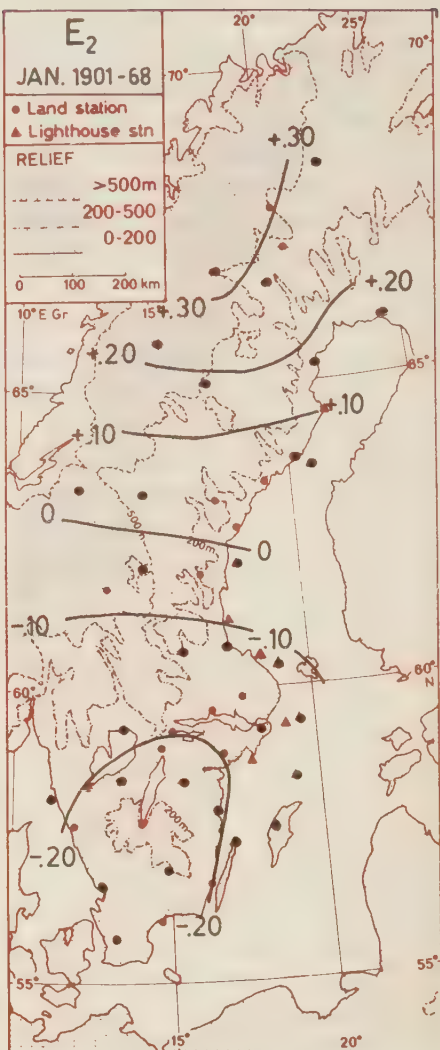


Fig. 3b: Geographical distribution of the components of the second eigenvector (the second Characteristic Pattern) for January.

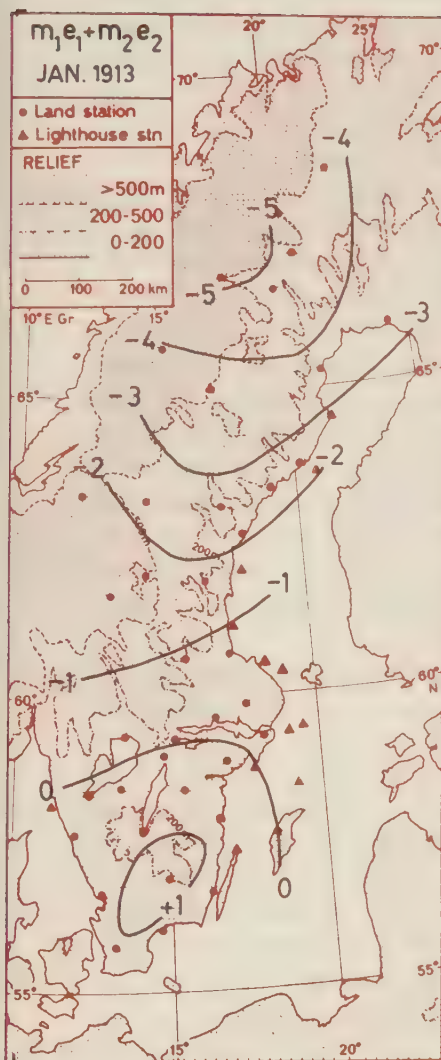


Fig. 4a: Approximation with two eigenvectors to the temperature anomalies during January 1913.

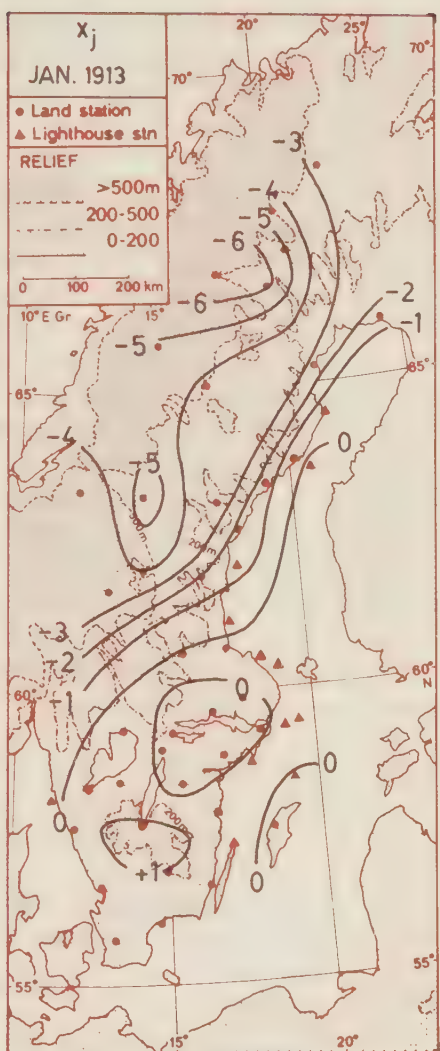


Fig. 4b: Observed temperature anomalies during January 1913.

For examining the similarity of the patterns it is convenient to regard each characteristic pattern as a 31-component vector and for each pair of them compute their inner or scalar product, i.e.

$$\sum_{j=1}^p e_{jr} e_{js}$$

Here  $r$  and  $s$  denote the number of the principal axis resp. For one and the same month this sum is  $= 1$  if  $r = s$  and  $= 0$  if  $r \neq s$ . This follows from the orthogonality of the patterns. In other words if the patterns are identical their scalar product is unity.

No such relationships hold if we take pairs from different months. Then we may regard the scalar product as an index giving the similarity between the patterns. Pairs having an inner product  $> 0.95$  are very similar and according to Grimmer (1963) an inner product of 0.7 implies similarity of a high order. In Table 1 the inner products with the January characteristic patterns are indicated by lines.

Table 1: Fractions of the total sample variance ( $V_k$ ) obtained from the first five Characteristic Patterns (C.P.)

Double lines indicate inner products  $\geq 0.98$  with the January C.P.

Single " " " "  $\geq 0.95$  " " " " "  
Broken " " " "  $\geq 0.90$  " " " " "

| C.P.<br>No | Jan | Feb | Mar | Apr | May  | Jun | Jul | Aug  | Sep  | Oct | Nov  | Dec |
|------------|-----|-----|-----|-----|------|-----|-----|------|------|-----|------|-----|
| 1          | =81 | =85 | =87 | -80 | -79  | -79 | -77 | -82  | -81  | =83 | =83  | =82 |
| 2          | =12 | =8  | -7  | --9 | ---7 | -12 | -13 | ---9 | ---9 | =11 | =8   | =11 |
| 3          | 2   | 2   | 2   | 3   | 4    | 5   | 3   | 2    | 2    | --2 | ---3 | 2   |
| 4          | 2   | 1   | 1   | 2   | 2    | 2   | 1   | 2    | 2    | 1   | 2    | 1   |
| 5          | 1   | 1   | 1   | 1   | 1    | 1   | 1   | 1    | 1    | 1   | 1    |     |

The two first patterns are nearly identical for all months. The components of the first characteristic vector are always negative, as the January pattern in Fig. 3a. The components of the second characteristic vector are negative over southern Sweden, positive over northern, and the zero isopleth goes approximately along latitude  $62^{\circ}\text{N}$ , see Fig. 3b.

Since the components of the first C.P.s (eigenvectors) have the same sign, they represent an anomaly having the same sign over the whole region. Remembering the large contribution to the total sample variance from this C.P., it asserts the fact that commonly the actual temperature anomaly has the same sign over the whole country. A contribution towards different signs of the actual anomalies over northern and southern Sweden is given by the second C.P.

For our samples these reproductions, using only the two first principal components, are very efficient since for every month about 90 % of the total sample variances is "explained". Thus we can satisfactorily describe the monthly temperature distribution over Sweden by using only the first and second principal components (multipliers) instead of the 31 observed values, i.e. we have reduced the data stock with more than 90%.

These two figures (the first and second multiplier) may be used for a numerical classification of the monthly temperature distribution. This method of classification is probably superior to any qualitative one.

#### B. Principal component analysis of the monthly temperature anomalies at Stockholm

Temperature observations have been regularly performed at Observatoriekullen in Stockholm since 1756. For Sweden this record is unique, since the station has never been removed. The material used here was the monthly temperatures from the period 1756 - 1967 ( $N = 212$ ). Since the year consists of twelve months we have  $p = 12$ . The analysis is performed in the same manner as in A and we have worked with the deviations from the individual (or climatological) monthly means, i.e.

$x_{ij}$  = deviation from the individual monthly mean during month  $i$ , year  $j$ .

The components of the first eigenvector are of one sign, reaching their highest absolute values during the colder months. The last feature is characteristic also for the second and third eigenvector and, though to a less degree, for the fourth and fifth. It is an expression of the fact that the winter temperatures have a much higher variance than the summer ones.

The principal components belonging to the first eigenvector are highly correlated with the annual temperatures ( $r = -0.84$ ). Thus a high negative value of the first principal component indicates a high annual temperature and a small temperature difference between summer and winter. The small annual range is caused by the fact that the components of the first eigenvector are close to zero during the warmer months and much below zero during the colder. Conversely, a high positive value of the first principal components indicates a low annual temperature and a large annual temperature range. Thus during the period investigated the warmer years have tended to be of a more maritime type than the colder ones. Most of the negative first principal components have occurred during the years after 1880.

Eigenvectors no.s 2, 3, 4 and 5 mainly represent different types of variations in temperature during the colder months. The absolute values of their components are then often high and the sign may change from month to month, see e.g. the December and January components of the second eigenvector.

Table 2: Fraction of the total sample variance ( $V_k$ ) obtained from the first five eigenvectors (EV) and their components.

| EV                                             | No    | 1     | 2     | 3     | 4     | 5     |
|------------------------------------------------|-------|-------|-------|-------|-------|-------|
|                                                | $V_k$ | 31.8  | 14.0  | 11.1  | 9.0   | 7.1   |
| C<br>O<br>M<br>P<br>O<br>N<br>E<br>N<br>T<br>S | Dec   | -.121 | -.558 | -.618 | .234  | -.159 |
|                                                | Jan   | -.463 | .524  | -.469 | -.496 | .068  |
|                                                | Feb   | -.672 | .081  | .232  | .348  | -.566 |
|                                                | Mar   | -.491 | -.201 | .235  | .182  | .684  |
|                                                | Apr   | -.217 | -.143 | .120  | -.107 | .250  |
|                                                | May   | -.100 | -.320 | .024  | -.321 | .030  |
|                                                | Jun   | -.028 | -.181 | .161  | -.325 | -.080 |
|                                                | Jul   | -.027 | -.191 | .245  | -.353 | -.075 |
|                                                | Aug   | -.057 | -.233 | .219  | -.273 | -.199 |
|                                                | Sep   | -.044 | -.123 | .073  | -.307 | -.192 |
|                                                | Oct   | -.077 | -.223 | -.013 | -.160 | -.137 |
|                                                | Nov   | -.092 | -.238 | -.368 | -.075 | .108  |

A similar analysis of the Central England temperatures during 1680 - 1963 was carried out by Craddock (1965). A comparison between the eigenvectors representing the Central England temperatures and those representing the Stockholm ones is of interest. We measure the similarity with their scalar products, i.e.

$$\sum_{j=1}^{12} e_{jr}^c e_{jt}^s$$

where

$$\begin{aligned} e_{jr}^c &= \text{component } j \text{ of eigenvector } r \text{ for Central England} \\ e_{jt}^s &= \text{ " " " " } t \text{ " Stockholm} \end{aligned}$$

The highest scalar products for combinations of the first four eigenvectors are:

$$\sum_j e_{j1}^c e_{j1}^s = -0.905$$

$$\sum_j e_{j2}^c e_{j3}^s = -0.786$$

$$\sum_j e_{j3}^c e_{j2}^s = 0.946$$

$$\sum_j e_{j4}^c e_{j4}^s = 0.807$$

All the components of the first eigenvector for Central England ( $e_{j1}^c$ ) have the same sign, just as in the Stockholm case ( $e_{j1}^s$ ). Remembering that the scalar product is unity if the vectors are identical, the high absolute values of the scalar products of these eigenvectors show that they are very similar. (We need not bother about the sign. It may be determined arbitrarily.) These high scalar products indicate similarities in the annual march of the temperature anomalies in England and Sweden.

The efficiency of the Central England representation was also similar to the Stockholm one. Craddock obtained the following results:

$$V_1 = 27.5, \quad V_2 = 15.2, \quad V_3 = 10.8, \quad V_4 = 10.8$$

Though the annual march of the temperature anomalies is very complicated, and the principal component representation is not very efficient, the analysis gives valuable information.

The purpose of the first meteorological works using principal component analyses, or in Lorenz's (1956) terminology, empirical orthogonal functions, was weather forecasting. Here the progress seems to have been slow.

Possibly principal component analyses in the future may be of use in locating predictors for forecasting purposes. At present, however, their main uses are descriptive and explanatory and as such they constitute an excellent tool for climatological studies.

Acknowledgements: I have had valuable discussions with Prof. G.H.Liljequist, Prof. T.Bergeron, Dr U.Högström, Dr. P.Seeger, Prof. H.Köhler and Dr.C.J.Östman. Prof. G.H.Liljequist has put part of the resources of the Meteorological Institute at my disposal. Prof. T.Bergeron has given me permission to use observations from Project Pluvius. Mrs A-S.Smedman has helped me to collect and check the temperature data from Sweden. Concerning both the Swedish temperature and precipitation data, senior meteorologist B.Eriksson at the Swedish Meteorological and Hydrological Institute has helped me to collect and check them. Much data extracting and checking has been performed by Mrs. B.Löfgren, Mr J.Bråholm, Mr. P.Herman and Mr. J.Lapins. Mr. Lapins is also responsible for most of the drawings. The typewriting has been performed by Mrs. M.Johnsson, Mrs. B.Löfgren and Mrs. E-M.Öhman. The programming work has been performed by Mr. J.Keijser and Mr. J.Urmi. Mrs. G.Hörner and Dr. R.Fenimore have corrected my English. Funds have been furnished by the Swedish Natural Science Council and Statens Råd för Byggnadsforskning. To all these persons and Institutes I express my sincere thanks.

#### References:

- Bergeron, T., 1930: Richtlinien einer dynamischen Meteorologie. Met. Zeitschr., Vol. 65.
- Bielke, S. and Georgii, H-W., 1968: Investigation of the incorporation of sulfur-dioxide into fog- and rain-drops. Tellus, Vol. 20, Nr 3.
- Craddock, J.M., 1965: A meteorological application of principal component analysis. The Statistician, Vol. 15.
- Liljequist, G.H., 1943: The severity of winters at Stockholm. Sveriges Met. och Hydr. Anst., Ser. Upps., Nr 46.
- Liljequist, G.H., 1950: On fluctuations of the summer mean temperature in Sweden. Sveriges Met. och Hydr. Inst., Medd. Ser. B, Nr 7.
- Liljequist, G.H., 1970: Klimatologi. Generalstabens Litografiska Anstalt.
- Lorenz, E.N., 1956: Empirical orthogonal functions and statistical weather forecasting. Mass. Inst. of Technology, Dept. of Met., Scientific Report No. 1, Statistical forecasting project.
- Morrison, D.F., 1967: Multivariate statistical methods. McGraw-Hill series in probability and statistics.
- Rodhe, H., 1969: Om uppehållstiden för antropogent svavel i atmosfären. Met. Inst., Stockholms Universitet, Rapport AG/1.
- Wallén, A., 1930: Climate of Sweden. Sveriges Met. och Hydr. Anst., Nr 279.
- Ångström, A., 1946: Sveriges Klimat. Generalstabens Litografiska Anstalt. First edition 1946, second 1958 and third (in print) 1970.



